Methods for collection and dealing with complex data sets
Rita Walton and Anubhav Bagley

Abstract: The Maricopa Association of Governments (MAG) is responsible for regional land use, transportation and air quality modeling for the Phoenix metropolitan area. MAG uses a variety of models and geographies in its planning process. MAG's socioeconomic and land use model, SAM-IM, requires a complex data structure not usually associated with GIS to model land use changes and then prepares input data for the transportation models. These activities require continuous collection and update of data resources at varied levels of geographies and time periods. This paper describes the methods and techniques used to effectively organize and utilize such data.

GIS at MAG

The Maricopa Association of Governments (MAG) is responsible for regional land use, transportation, and air quality modeling for the Phoenix metropolitan area. Like other regional governments with similar MPO responsibilities, MAG has been seeking to improve the performance of these models, thereby making the regional planning process more effective. GIS has played a major role at MAG for the past twelve years. This technology has significantly enhanced MAG's ability to manage dozens of planning databases: databases describing census statistics, employment inventories, land use, general plans, parcels, building permits, and highway infrastructure – all sharing one common characteristic -- they are essentially spatial in nature. GIS helps MAG assume an important responsibility in the region – the role of a regional information agency for public planning.

GIS has capabilities that run far beyond just being a platform for planning and mapping. In recent years, MAG has been developing new classes of planning models that run directly on top of GIS. This new class of models works directly from GIS databases. More importantly, these models tap a wide variety of powerful spatial analysis methods that are found only in GIS and cannot be even remotely matched by older custom-written programs developed over the last three decades. In the long term, this new class of GIS-based models offers the potential for completely replacing older, non-spatial planning models.

There are a wide variety of ongoing data collection efforts ongoing at MAG. These are necessary to feed the complex land use and socioeconomic models as well as the transportation and air quality models. This paper discusses the data collection processes that are being used and also the nature and structure of the databases that attempt to quantify qualitative variables.

Background on SAM-IM

MAG has developed a rule-based urban growth model called SAM-IM (Sub-Area Allocation Model – Information Manager). It simulates both short-term and long-term urbanization of a region by reacting to any set of factors and conditions that the planner wishes to express. The model is completely embedded in a Geographic Information System (GIS) – it runs on ArcView GIS using the Spatial Analyst extension. The concept of the model is entirely GIS oriented – all of the data that drives the model, whether it be existing land use distributions, future market conditions, adopted planned land use, developments already approved and underway, or local land conditions, are expressed geographically, in the form of ArcView shape files. Anything that can be expressed geographically can be taken into consideration in the model. The Information

Manager part of SAM is the most important with respect to the wealth of data MAG collects to augment the GIS coverages.

There are several land use themes that drive the logic in SAM-IM. These are:

- ❑ Existing Land Use
- ❑ Planned Land Use (General Plan)
- ❑ Active Development Projects
- ❑ Planned Development Projects
- ❑ Proposed Development Projects
- ❑ Redevelopment Projects

In addition to these "input" land use themes, SAM-IM itself creates new land use databases, including one identifying growth by land use. By cumulating the growth to the Existing Land Use coverage, an Existing Land Use coverage is created for each forecast year. These are identified as Forecast Land Use coverages.

SAM-IM also ensures that special population group features are depicted in the "existing land use" theme. Therefore, a SAM-IM "compliant" existing land use theme is constructed through an overlay of point features appearing in special population group themes with the underlying existing land use polygons in order to build a comprehensive existing land use theme that includes special population group features as well as underlying population and employment densities.

Currently the special population group themes include:

- ❑ Seasonal Population Group:
  - ✓ Mobile Homes and RV Parks (Point features)
  - ✓ "Official" Retirement Communities (Polygon features)
- ❑ Transient Population Group:
  - ✓ Motels, Hotels, Resorts (Point features)
- ❑ Group Quarters Population Group:
  - ✓ Nursing Homes (Point features)
  - ✓ Prisons (Point features)
  - ✓ Dormitories (Point features)
  - ✓ Military Barracks (Point features)

Other variables considered by SAM include:
  - ✓ High School and Post High School Enrollment (Point features)
  - ✓ Airport Enplanements (Point features)
  - ✓ Floor Area Ratios by Land use type (Polygon features)
  - ✓ Residential Vacancy Rates by unit type (Polygon features)
  - ✓ Persons per Household by unit type (Polygon features)
  - ✓ Employees per 1000 sq. ft by land use type(Polygon features)

SAM-IM Functions

Predictions of future land absorption in SAM-IM are predicated on overlaying the various input land covers. Land depicted in the existing land use theme is superimposed on land depicted in the plan theme in order to identify sites that are both vacant (existing land use) and dedicated to an appropriate use (plan), and therefore eligible to absorb growth of a certain type. Projects that appear in the active developments theme is another important source of information about where growth will occur – these polygons are always given priority consideration in the forecast allocation.

Also, SAM-IM is a quantitative forecasting model – it is not only dealing with allocating "built uses" but it is also addressing housing and employment densities associated with those "built uses" (so that Traffic Analysis Zone (TAZ) data sets can be created to drive the EMME/2 transportation model). SAM-IM works best when it has an "existing land use theme" that carries actual socioeconomic characteristics for the polygons – actual dwelling unit densities for residential polygons and employment statistics for the commercial land use polygons.

SAM-IM forecasts growth in population and employment. MAG distinguishes several types of employment spatially: retail, office, industrial, public, and "other". Also, there are different types of population of interest: resident population, transient population (visitors), seasonal population, and group quarters. Residential population is spatially allocated in the form of dwelling units; transient population is allocated spatially in the form of hotel and motel rooms; seasonal populations are allocated spatially in terms of mobile home and RV sites; and group quarters populations are allocated in terms of the root type of group quarters population group in question: military barracks, dormitories, nursing homes, and prisons.

The groups that are allocated spatially in SAM-IM are known collectively as allocation sectors. The land use databases must characterize all land in the study area as to the land use, the type of sector (residents or employment), and the quantity (or density). SAM-IM is programmable and can allocate, from regional, countywide, or RAZ-level forecasts, any set of allocation sectors. Currently, there are 14 such sectors, as shown in Table 1.

SAM-IM allocates growth associated with each of the 14 sectors to vacant and developable land in the modeling area. Each cell in the resulting growth grid is coded according to a an existing land use category along with estimates of sectors and densities associated with the allocation sector that is projected to occupy it.

Table 1: SAM-IM Allocation Sectors

Category	Allocation Sector	Allocation Units	Description
Resident Population	"RSF"	Dus	Allocation of single-family resident (non-special population) dwelling units
	"RMF"	Dus	Allocation of multi-family resident (non-special population) dwelling units
Transient Population	"Mot"	Rooms	Allocation of motel/hotel/resort room from which transient populations are later derived
Special Population	"MH"	Spaces/Units	Mobile home and RV spaces in formal "parks" from which resident and seasonal populations are later derived
	"Rtmt"	Dus	Allocation of dwelling units in official "retirement communities", from which resident and seasonal populations are later derived.
Group Quarters	"NrsH"	Beds	Allocation of nursing home population
	"Mil"	Beds	Military
	"Dorm"	Beds	Post-HS Dormitories
	"Psn"	Cells	Prisons/jails
Employment	"Retl"	Employees	Retail Employment
	"Off"		Office
	"Ind"		Industrial
	"Pub"		Public
	"Oth"		Other
	"Home"		"Work-at-home" employees

Data Structure and Issues

In preparation for the next set of socioeconomic projections to be generated in 2002 MAG has spearheaded a major data collection effort that began in 1999. As part of this MAG GIS and Database Enhancement Project four consultants have been working with MAG to collect and collate all the data required for the models. The data being collected includes an updated Street Centerline file for Maricopa County and also the areas in Pinal County that are within the MAG modeling region. Existing land use, General Plan land use, and Development areas are the other major GIS coverages being developed. An employment database documenting all employers with 5 or more employees, a building inventory, and special population inventories are also being developed as part of this task. Major assumptions are also being revised.

Many data structure issues and concepts have contributed to this effort:

Land Use Categories
The current land use data used a 25-category land use code. This code did not completely meet the demands of this fast expanding area or the enhanced MAG land use models. Table 2 lists the new 46-category hierarchical land use codes that were developed using input from all MAG member agencies. The new categories were designed with flexibility for future expansion and changes.

Table 2: MAG Land Use Coding Scheme

Code	Land Use:	Code	Land Use:
100	General Residential	500	General Employment
110	Rural Residential	510	Tourist and Visitor Accommodations
120	Estate Residential	520	Educational
130	Large Lot Residential (SF)	530	Institutional
140	Medium Lot Residential (SF)	540	Cemeteries
150	Small Lot Residential (SF)	550	Public Facilities
160	Very Small Lot Residential (SF)	560	Special Events
170	Medium Density Residential (MF)	570	Other Employment (low)
180	High Density Residential (MF)	580	Other Employment (medium)
190	Very High Density Residential (MF)	590	Other Employment (high)
200	General Commercial	600	General Transportation
210	Specialty Commercial	610	Transportation
220	Neighborhood Commercial	620	Airports
230	Community Commercial	700	General Open Space
240	Regional Commercial	710	Active Open Space
250	Super-Regional Commercial	720	Golf courses
300	General Industrial	730	Passive Open Space
310	Warehouse/Distribution Centers	740	Water
320	Industrial	750	Agriculture
400	Office General	800	Multiple Use General
410	Office Low Rise	810	Business Park
420	Office Mid Rise	820	Mixed Use
430	Office High Rise	900	Vacant (existing land use database only)

Land Use Data Model

The basic premises for the land use data model are:

❑ that the existing land use theme, the general plan theme, and the developments theme are ALL polygon feature shape files.

❑ that six of the seven special population group coverages (prisons, nursing homes, dormitories, military barracks, motels, mobile home sites) will be point feature shape files. These point features will later be appended to the existing land use polygons through an automated procedure, and attributes associated with these special population groups will be migrated to the existing land use theme.

❑ that the special population group and retirement communities coverage will be a polygon feature shape file.

Linked Land Use Data Model

The new MAG data structure allows for multiple entries with a parent-child relationship. There is a one-to-many relationship between the parent file (MASTER) and the child file (CONTENT). In the basic land use data model all land use themes therefore consist of a minimum:

❑ **Master File:** A basic polygon shape file with five attributes: (1) a unique polygon identifier (the database key), (2) the primary MAG land use code for the polygon, (3) the area of the polygon (in acres), (4) the jurisdiction (MPA) in which the polygon is located, and (5) the jurisdiction (MPA) land use code.

❑ **Content Table:** Associated with each land use theme is a second table, also a dBase table, which we call a content table. The data model supports one-to-many relations between the polygons that appear in the Master File and the descriptors of the contents of the polygons that appear in the content table.

In other words, there can be more than one record that describes the contents of a land use polygon, for each polygon. One land use polygon therefore can identify residential dwelling units, employment of various types, and special population group features such as motels, mobile homes, etc. This is very important as mixed-use categories are being widely utilized by planners in the General Plans.
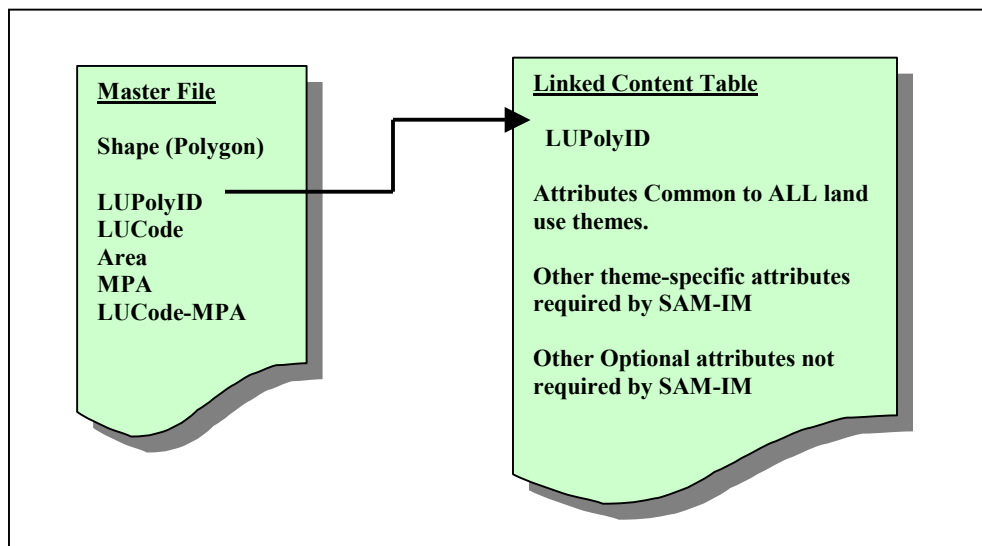
ArcView can support one-to-many relations automatically with a "table-link" operation. The implication of this data model is simply that there are no constraints – any land use polygon can contain any land use or any combination of land uses. It is totally flexible.

The Content Table

The ArcView dBase file describing the contents of the land uses in the polygons will have:

1. *Common Attributes for ALL Land Covers:* a set of attributes that apply to ALL land use covers. One of these attributes is the unique polygon number that appears in the Master File itself and therefore is the key on which a "table-link" operation is performed.

2. *Theme-Specific Attributes for Individual Land Covers Needed by SAM-IM:* a set of theme-specific attributes that are minimally required by SAM-IM. For example, the developments theme has attributes specifically required of developments that do not appear in the existing land use theme.

3. *Other Theme-Specific Attributes:* other theme-specific attributes that are of general interest even though they are not used specifically by SAM-IM.

Figure 1:The Linked Land Use Data Model

Master File

Shape (Polygon)

LUPolyID
LUCode
Area
MPA
LUCode-MPA

Linked Content Table

LUPolyID

Attributes Common to ALL land use themes.

Other theme-specific attributes required by SAM-IM

Other Optional attributes not required by SAM-IM

Applicability

An important concept integral to the programming of SAM-IM is that the same database structure is used for all on-screen editors, GUI's, Avenue scripts, and allocation routines, no matter which land use theme is represented. In addition, the growth and forecast land use themes that are created by SAM-IM conform to this model.

Land Use Database Structure

Existing Land Use Theme

Table 3 describes the files associated with the Existing Land Use and the attribute fields associated with this theme. Further explanations are in Appendix A.

Table 3: Existing Land Use Theme

Table	Field Name	Data Type	Field Width	Decimals	Description
Master File	**LUPolyID**	I	8	0	Unique polygon identifier (key)
	LUCode	I	3	0	Land use code: use MAG coding dictionary. This is the primary land use code, for example 110 for Rural Residential and 820 for Mixed use.
	Acres	F	9	2	Polygon area (in acres)
	MPA	C	2	0	MPA: use MAG list of MPAs.
	LUCodeMPA	I	3	0	Land use code as defined by the MPA if available. A table of land use codes for each MPA will be created for this.
	LastUpdate	D			Date of last update (by technician)
	EffectiveDate	D			Date when the General Plan was last modified by city/town action
Content File	**LUPolyID**	I	8	0	Unique polygon identifier: this identifier matches the polygon identifier in the Master File.
	LUCode2	I	3	0	Secondary Land use code: use MAG coding dictionary. This cannot be a multiple use category.
	Sector	C	4	0	Sector: see abbreviations listed for each allocation sector in table 2.
	Pct	F	8	4	Percentage of the land area covered by this record. This is important (for multiple use polygons) because the density "value" field below will quote values for net density for the portion of the polygon covered by the record, not the gross density across the area of the entire polygon.
	Units	C	1	0	"d" or "q" A "d" means that the value field represents density. A "q" means that the value field represents an absolute quantity (e.g., 400 Dus, not 400 Dus per acre)
	TgtMAG	F	12	4	Density Value or Quantity Value (e.g., number of units), depending on the coding of the "units" field above. For existing land use, this is the density or quantity value that is actually built.

General Plan Theme

The table structure and attribute fields associated with the General Plan Land Use theme are in Table 4. Further notes on the structure are in Appendix A.

Table 4: General Plan Land Use Theme

Table	Field Name	Data Type	Field Width	Decimals	Description
Master File	LUPolyID	I	8	0	Unique polygon identifier (key)
	LUCode	I	3	0	Land use code: use MAG coding dictionary. This is the primary land use code, for example 110 for Rural Residential and 820 for Mixed use.
	MPA	C	2	0	MPA: use MAG list of MPAs.
	LUCodeMPA	C	5	0	Land use code as defined by the MPA. A table of land use codes for each MPA will be created for this.
	Acres	F	9	2	Polygon area (in acres)
	LastUpdate	D			Date of last update (by technician)
	EffectiveDate	D			Date when the General Plan was last modified by city/town action
Content File	LUPolyID	I	8	0	Unique polygon identifier: this identifier matches the polygon identifier in the Master File.
	LUCode2	I	3	0	Secondary Land use code: use MAG coding dictionary. This cannot be a multiple use category.
	Sector	C	4	0	Sector: see abbreviations listed for each allocation sector in table 2.
	Pct	F	8	4	Percentage of the land area covered by this record. This is important (for multiple use polygons) because the density "value" field below will quote values for net density for the portion of the polygon covered by the record, not the gross density across the area of the entire polygon.
	Units	C	1	0	"d" or "q" A "d" means that the value field represents density. A "q" means that the value field represents an absolute quantity (e.g., 400 Dus, not 400 Dus per acre)
	TgtMAG	F	12	4	For general Plan, this will be the MAG defined Density Value or Quantity Value (e.g., number of units), depending on the coding of the "units" field above. This is the "nominal allocation density" at which SAM-IM will absorb development.

Table	Field Name	Data Type	Field Width	Decimals	Description
	MinMAG	F	12	4	This is the minimum density that is expected for this general plan polygon. This is defined by MAG by land use. SAM-IM will attempt to allocate according to the "value" field, unless it cannot do so (due to insufficient growth)
	MaxMAG	F	12	4	This is the maximum density that is expected for this general plan polygon. This is defined by MAG by land use. SAM-IM will attempt to allocate according to the densities coded in the "value" field, unless there are reasons for it to increase densities (due to abnormally high growth from DRAM/EMPAL).
	TgtMPA	F	12	4	MPA defined Density Value or Quantity Value (e.g., number of units), depending on the coding of the "units" field above.
	MinMPA	F	12	4	This is the MPA defined minimum density that is expected for this general plan polygon.
	MaxMPA	F	12	4	This is the MPA defined maximum density that is expected for this general plan polygon.

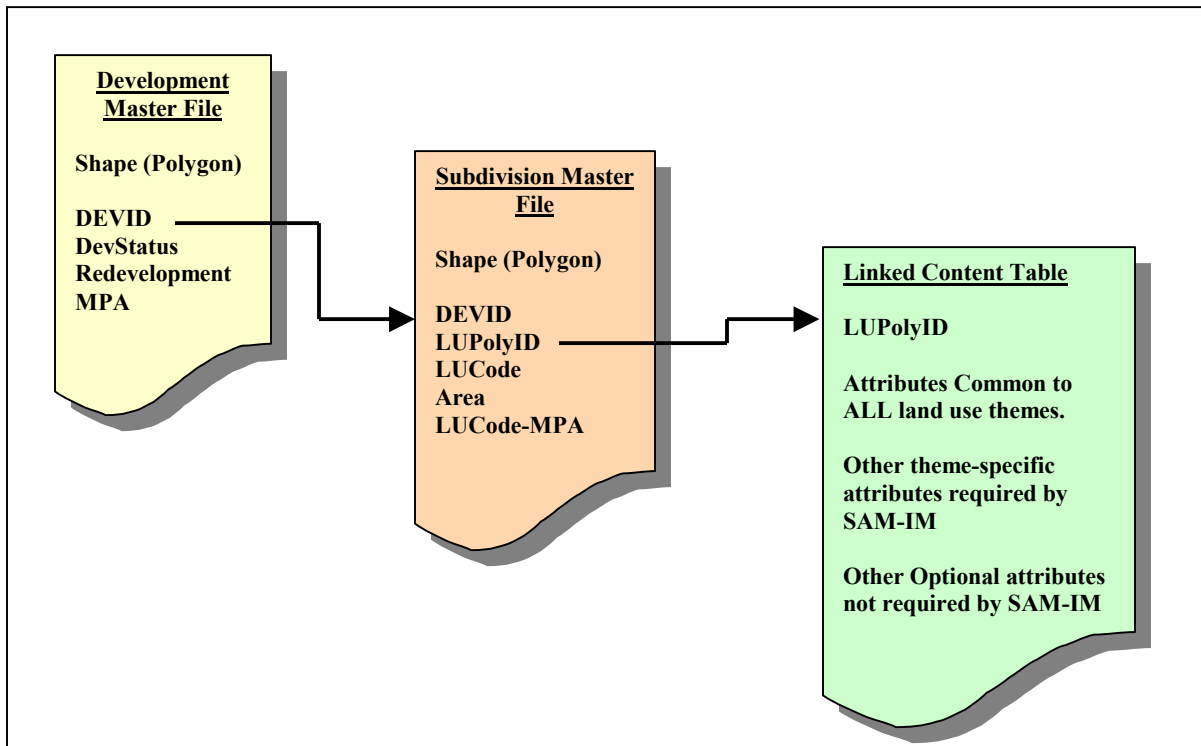## Developments Theme (Active, Planned, Proposed & Redevelopment)

This section deals with the four development themes, namely Active, Planned, Proposed, and Redevelopment.

- **Active Developments :** represent a compilation of projects that are already under construction or are expected to be under construction within the modeling horizon (typically 5 years). SAM-IM allocates to these projects according to the concept of "development velocities" in a separate allocation pass. The active developments theme contains attributes that describe the expected beginning and completion dates associated with the project along with the age of the project.

- **Planned Developments:** represent a compilation of projects that have at least been through the Planning and Zoning approval process but have not begun construction. SAM-IM does NOT allocate to these projects automatically; instead it favors them by incorporating a "probability" score associated with them in the site suitability evaluation.

- **Proposed Developments:** represent projects for which land use plans are available but have not yet been through any formal approval process. These projects will also have probability scores associated with them, usually lower than those for Planned Developments.

- **Redevelopment Projects:** represent existing built areas that are being planned for redevelopment. An example of a project can be a downtown redevelopment plan. SAM-IM uses these projects to identify the new land use and also estimate the population and employment being relocated due to the project.

Developments Theme Structure

The database structure for all of the Development themes is different from the Existing and General Plan land use as this has a three-tier structure. The first tier primarily contains information about the complete development. The second and third tiers are similar to the Master file (Subdivision Master File) information and the Content file. Two spatial coverages are created – the first with all specific development projects and the second containing the subdivisions composing the projects. Generally, the total geographic area covered by the two is the same.

Figure 2:The Linked Development Theme Structure

The attribute fields associated with the Development themes, which includes Active, Planned, Proposed, and Redevelopment projects, are identified in Table 5. Further notes are in Appendix A.

Table 5: Development Theme (Active, Planned, Proposed & Redevelopment Projects)

Table	Field Name	Data Type	Field Width	Decimals	Description
Development Master File	**DevID**	I	8	0	Unique polygon identifier for entire development (key)
	DevID1996	I	4		ID from 1996 study
	MPA	C	2	0	MPA: use MAG list of MPAs.
	DevName	C	50		Common Name of Entire Development
	DevStatus	C	4	0	This field provides the status of the entire project i.e Active Development ("Actv"), Planned Development ("Plnd"), Proposed Development ("Prop") or Complete ("Comp").
	RedevProj	C	1	0	Is this a Redevelopment Project (Y/N)
	DevArea	F	9	2	Polygon area for the entire development (in acres)
	DevStage	C	4	0	Where the Development Proposal is in the Approval Process (codes to come)
	ZoningID	C	10	0	Original Zoning Case Number for Entire Area
	LastUpdate	D			Date of last update
	EffectiveDate	D			Date when the plan was last modified by city/town action. This provides a benchmark date for the last action on the project.
	DevStartYear	I	4		Estimated start year of Development for Entire Area
	DevStartQtr	I	1		Estimated starting quarter of Development for Entire Area
	DevEndYear	I	4		Estimated completion year of Development for Entire Area
	DevEndQtr	I	1		Estimated ending quarter of Development for Entire Area.
	SitePlan	C	1		Is there a current site plan (Yes/No).
	Comments	C	255		Additional Comments
Subdivision Master File	**DevID**	I	8	0	Unique polygon identifier for entire development: this identifier matches the polygon identifier in the Development Master File.
	LUPolyID	I	8	0	Unique polygon identifier for each subdivision in the development (key).
	LUCode	I	3	0	Land use code: use MAG coding dictionary. This is the primary land use code, for example 110 for Rural Residential and 820 for Mixed use.

Table	Field Name	Data Type	Field Width	Decimals	Description
	LUCodeMPA	C	5	0	Land use code as defined by the MPA. A table of land use codes for each MPA will be created for this.
	SubdivName	C	50		Name of the project on this subdivision
	SubdivStatus	C	4		This field provides the status of the subdivision i.e Active Development ("Actv"), Planned Development ("Plnd"), Proposed Development ("Prop") or Complete ("Comp").
	RedevSubdiv	C	1	0	Is this Redevelopment (Y/N)
	SubdivArea	F	9	2	Polygon area for the subdivision (in acres)
	ZoningID	C	10		Original Zoning Case Number for this part of the development
	MeyersDistrict	C	6		
	MeyersPage	C	6		
	MeyersID	C	6		
	Comments	C	255		Additional Comments
Content File	LUPolyID	I	8	0	Unique polygon identifier for each subdivision in the development: this identifier matches the polygon identifier in the Subdivision Master File.
	LUCode2	I	3	0	Secondary Land use code: use MAG coding dictionary. This cannot be a multiple use category.
	Sector	C	4	0	Sector: see abbreviations listed for each allocation sector in table 2.
	Pct	F	8	4	Percentage of the land area covered by this record. This is important (for multiple use polygons) because the density "value" field below will quote values for net density for the portion of the polygon covered by the record, not the gross density across the area of the entire polygon.
	Units	C	1	0	"d" or "q" flag. A "d" means that the value field represents density. A "q" means that the value field represents an absolute quantity (e.g., 400 Dus, not 400 Dus per acre)
	TgtMAG	F	12	4	MAG defined Density Value or Quantity Value (e.g., number of units), depending on the coding of the "units" field above.
	MinMAG	F	12	4	This is the minimum density that is expected for this land use. This is defined by MAG by land use. SAM-IM will attempt to allocate according to the "value" field, unless it can not do so (due to insufficient growth)

Table	Field Name	Data Type	Field Width	Decimals	Description
	MaxMAG	F	12	4	This is the maximum density that is expected for this land use. This is defined by MAG by land use. SAM-IM will attempt to allocate according to the densities coded in the "value" field, unless there are reasons for it to increase densities (due to abnormally high growth from DRAM/EMPAL).
	TgtMPA	F	12	4	MPA defined Density Value or Quantity Value (e.g., number of units), depending on the coding of the "units" field above.
	MinMPA	**F**	**12**	**4**	This is the MPA defined minimum density that is expected for this land use.
	MaxMPA	F	12	4	This is the MPA defined maximum density that is expected for this land use.
	BuiltAcres	F	9	2	Total acres developed
	BuiltAcres2000	F	9	2	Total acres developed as of July, 2000
	TotalSF	I	8		Total leasable Square Footage of non-residential (Res & Open Space: 0)
	BuiltSF	I	8		Built Square Footage of non-residential (Res & Open Space: 0)
	BuiltSF2000	I	8		Built sq. ft. July 2000
	TotalUnits	I	6		Total number of residential units to be built in this polygon (Non-res: 0)
	BuiltUnits	I	6		Residential Units Built (Non-res: 0)
	BuiltUnits2000	I	6		Units built July 2000
	StartYear	I	4		Estimated start year of land use development. The year in which allocations for the development should begin. This value MUST be no less than the start year for the entire project.
	StartQtr	I	1		Estimated starting quarter of land use development.
	EndYear	I	4		Estimated completion year for this land use. The year in which allocations for the development should be completed. If known, then the allocation will be staggered between the beginning year and the ending year. If left blank, then the allocation will be "completed" according to the development velocity curves.
	EndQtr	I	1		Estimated ending quarter of land use development.
	Prob	I	3	0	Project likelihood score 0..100 (100=absolutely certain). This score reflects the opinion of the planner as to the likelihood that this project will be built.

Looking Forward

As mentioned earlier, the new databases and GIS themes along with the enhanced SAM-IM model will be used to prepare the next set of socioeconomic projections for Maricopa County during 2002. For the next phase of work a number of data enhancements are planned. A few of these are essential for the projections, while others are more long range. These enhancements are:

Data Consistency Project: In the next few months MAG will have a large quantity of new data from a number of sources. This new data not only includes the land use databases and GIS themes but also the detailed information from Census 2000. A project is planned to ensure consistency between all of these varying data sets. The project will enhance the relationships between the spatial and attribute data. It is essential to have an existing land use theme that correctly reflects the population and employment numbers from Census 2000. This is necessary to make sure that the base densities being used for both employment and residential uses are correct. Similarly, checks are needed for all of the inventory databases such as Hotels and Motels, to ensure they are located on appropriate land uses.

Data Maintenance: Since extensive modeling efforts are based upon the data at MAG, it is essential to maintain the currency of this data. The consultants involved in the GIS and Database Update Project will be suggesting the methods necessary to maintain the data. It is planned to develop a GIS network of all the 27 MAG member agencies. The use of Arc-SDE is being considered to enable the sharing and maintenance of all the databases.

Other Data Characteristics: Land use is an extremely complex dataset. To effectively model the socioeconomic changes in an area it is necessary to accurately reflect the existing situation. In an area like Phoenix, where there is a large influx of people during winter months, it is important to understand the seasonal variations in land use. Also, there is a growing trend towards 24-hour work cycles especially in the service industry. These trends can dramatically change the resultant employment densities. Dealing with a few of these complex but growing issues will be necessary in the coming years.

Appendix A

Data Structure Notes

1. **LUPolyID** is the key to the record and therefore must be unique, even though it is arbitrary. Once established, the identifier must be maintained since it is now being used by the Content table to link on.

2. **LU Code** is MAG's official dictionary of land use codes as shown in Table 2.

3. **LUCode2** is the secondary land use code for each polygon. Generally, for uses other than multiple use this will be the same as LUCode. This cannot be one of the 800 series of the land use codes shown in Table 2.

4. The data model allows for multiple use designations. A multiple use polygon **MUST** be coded with multiple records in the CONTENT TABLE to show the mix of allocation sectors that can be allocated to it, their densities, and the percentage of land cover of the polygon that can be allocated.

5. **DevArea** and **SubdivArea** refer to the area of the polygon, expressed in acres.

6. **Sector** refers to the content of the land use polygon, expressed in terms of one of the 14 allocation sector types. Use the 4-character designator that appears in Table 1, so "Retl" refers to retail employment, etc.

7. **Pct** refers to the percentage (by land area) of the entire polygon that is occupied by the sector reported on tshis record, reported in the form nnn.nnnn, as in 100.0000 for 100%. All existing land use polygons should be coded this way. The field is especially important when coding multiple use polygons.

8. **Units** refer to how the "value" field is expressed – in terms of density (d) or in terms of absolute quantities (q). For the existing land uses, the appropriate expression is usually in terms of densities, as in "3.5 Dwelling units per acre". However, for other allocation sectors it may be much more convenient to express the magnitude of development in terms of absolute numerical quantities, as in "40 motel rooms".

9. Note that there is no field that defines the actual units that are being reported – that is whether "dormitories" quantities are reported in terms of "beds" or in terms of "rooms". SAM-IM actually permits anything, provided it is consistent across the model. The rule is that the units reported are implied by the sector, as defined in Table 1.

10. **TgtMAG** refers to the density or the quantity for the <u>Net Area</u> as signified by **Pct**. For example, a polygon of 10 acres having a multiple use designation with 60% Residential use (5 units per acre) and 40% Retail (50 employees) will mean a total of 30 residential units (60% * 10 acres *5 units/acre) and 50 retail employees (quantity for the use in the polygon).

11. **TgtMAG** for existing land use refers to the density or quantity that is already built.

12. **DevStatus** and **SubDivStatus** fields respectively provide the status of the entire project and subdivision. The values for this field can be Active Development (**Actv**), Planned Development (**Plnd**), Proposed Development (**Prop**), or Complete (**Comp**).

13. **RedevProj** is a Yes/No flag to mark Redevelopment Projects. **RedevSubdiv** in the Subdivision Master File is a flag to mark redevelopment areas within an Active/Planned or Proposed project.

14. The fields (1) **DevStartYear**, (2) **DevEndYear** are meant to establish the actual beginning and completion for the entire project and not for individual subdivisions.

15. **Prob** represents the project likelihood score. This is 100 for active projects.